On extracting common random bits from correlated sources

Andrej Bogdanov* Elchanan Mossel†

Abstract

Suppose Alice and Bob receive strings of unbiased independent but noisy bits from some random source. They wish to use their respective strings to extract a common sequence of random bits with high probability but without communicating. How many such bits can they extract? The trivial strategy of outputting the first k bits yields an agreement probability of $(1-\varepsilon)^k < 2^{-1.44k\varepsilon}$, where ε is the amount of noise. We show that no strategy can achieve agreement probability better than $2^{-k\varepsilon/(1-\varepsilon)}$.

On the other hand, we show that when $k \geq 10 + 2(1-\varepsilon)/\varepsilon$, there exists a strategy which achieves an agreement probability of $0.003(k\varepsilon)^{-1/2} \cdot 2^{-k\varepsilon/(1-\varepsilon)}$.

1 Introduction

Let x and y be strings in $\{0,1\}^n$ generated according to the following random process. First, each bit x_i of x is chosen independently at random from $\{0,1\}$. Then each bit y_i of y is independently set to equal x_i with probability $1-\varepsilon$ and $1-x_i$ with probability ε (the latter possibility indicates that x_i is corrupted). Suppose that Alice and Bob now want to agree on a common random string with probability at least, say, $1/2$. One possible protocol is for both of them to output the first $O(1/\epsilon)$ bits of their respective inputs. We show that no protocol can do better up to the constant factor. On the other hand we show that this gain by a constant factor can be achieved for certain values of the parameters.

This scenario relates to the problem of extracting a unique identification (ID) string from process variations. Several works have proposed hardware-based procedures for extracting a unique, uniformly random identifying string from a digital circuit of a given type [LLG+05, SHO08, YLH+09]. It has been proposed that such strings can be used for authentication and secret key generation of low-power devices such as RFIDs [LLG+05, SD07].

*Department of Computer Science and Engineering and Institute for Theoretical Science and Communications, Chinese University of Hong Kong. Email: andrejb@cse.cuhk.edu.hk. Supported by RGC GRF grant 2150617.

†U.C. Berkeley and Weizmann Institute. E-mail: mossel@stat.berkeley.edu. Supported by NSF Career award DMS-0548249 and Israeli Science Foundation grant 1300/08. Supported by Minerva foundation with funding from the Federal German Ministry for Education and Research

However, such procedures are prone to noise: Different instantiations of the procedure may produce slightly different answers. Can the agreement probability in any pair of instantiations be improved algorithmically while maintaining the uniform distribution of the ID string? Our work addresses this question when the noise is random and independent across the bits. We note that in applications, the noise can be handled using other methods, for example by incorporating noise tolerance at the receiver end.

The case where the goal of the two parties is to extract a single bit was studied independently a number of times. It is known that in this case the optimal protocol is for the two parties to use the first bit. See [Yan07] for references and for studying the problem of extracting one bit from two correlated sequences with different correlation structures.

In [MO05, MOR+06] a related question is studied: If m parties receive noisy versions of a common random string, where the noise of each party is independent, what is the strategy for the m parties that maximizes the probability that the parties agree on a *single* random bit of output without communicating? [MO05] shows that for large m using the majority functions on all bits is superior to using a single bit and [MOR+06] uses hyper contractive inequalities to show that for large m, majority is close to being optimal.

The optimality of the single bit protocol for two parties and extraction of one bit implies that if the goal of the two parties is to maximize the *expected* number of bits they agree on, given that they output k bits, they cannot do better than output the first k bits. However, this analysis leaves open the possibility that there exist a strategy where the two parties may be able to agree on *all* the bits with probability as large as $1-\varepsilon$.

We prove that this is not the case: The probability of agreement can be at most $2^{-k\varepsilon/(1-\varepsilon)}$. In the trivial strategy, where each party outputs its first k bits, the probability of agreement is $(1-\varepsilon)^k$. Figure 1 shows the ratio between the number of bits allowed by our upper bound and the performance of the trivial strategy, for any fixed agreement probability.

On the other hand, when the probability of agreement is sufficiently small, an improvement over the trivial strategy is possible: When $k \geq 10 + 2(1-\varepsilon)/\varepsilon$, there exists a protocol which achieves an agreement probability of $0.003(k\varepsilon)^{-1/2} \cdot 2^{-k\varepsilon/(1-\varepsilon)}$.

Our protocol is asymptotically almost optimal in the following sense. Suppose we want to achieve a fixed but sufficiently small agreement probability p. Our upper bound shows that if the trivial protocol extracts k bits, then no protocol can extract more than $(1/\ln 2)k$ bits. Our protocol can extract $(1/\ln 2 - \delta)k$ bits for any constant $\delta > 0$, as long as $\varepsilon = \varepsilon(\delta)$ is sufficiently small.

Gács and Körner [GK72] and Witsenhausen [Wit75] show that it is impossible for Alice and Bob to extract $\Omega(n)$ common random bits with probability $1 - o(1)$ for any finite distribution (x_i, y_i), unless x_i and y_i share common randomness. Our work applies to a specific (natural) distribution (x_i, y_i), but yields much sharper bounds. Maurer [Mau93] and Ahlswede and Csiszàr [AC93] consider a different model where Alice and Bob can communicate, but eavesdroppers are present and the common random string must remain secret. In this model, it is sometimes possible to achieve better agreement.

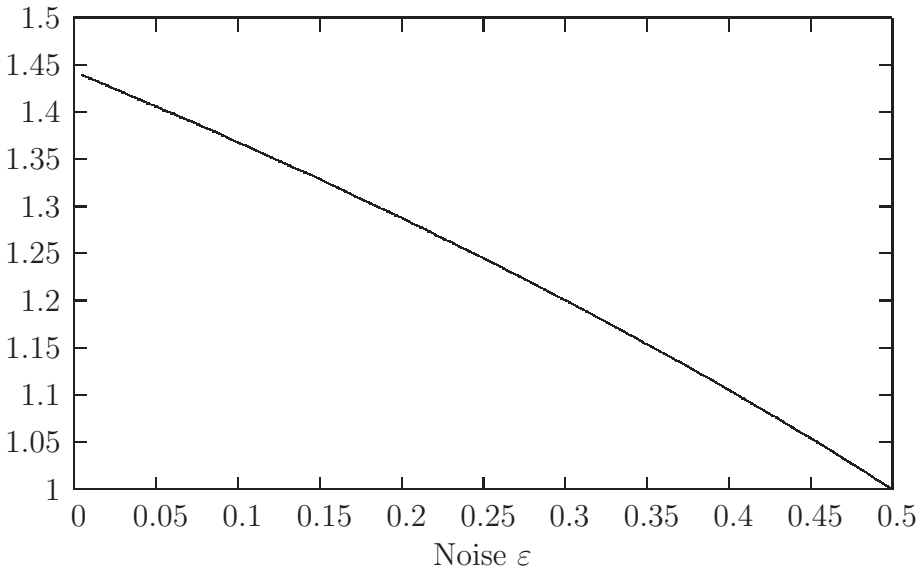

Figure 1: An upper bound on the factor beyond which the trivial protocol cannot be outperformed in terms of number of extracted bits (for any probability of agreement).

Notation Throughout the paper, we use n to denote the length of the correlated strings x and y available to Alice and Bob, k for the number of bits in their output, and ε for the noise. The inputs $x = x_1 \dots x_n$ and $y = y_1 \dots y_n$, $x_i, y_i \in \{0,1\}$ are chosen from the following distribution $(x,y)_\varepsilon$: Each pair $x_i y_i$ is independent of all the other pairs and takes the values 00, 11 with probability $(1-\varepsilon)/2$ each and the values 01, 10 with probability $\varepsilon/2$ each.

2 The upper bound

Consider a protocol where Alice and Bob produce k uniform bits of output. Such a protocol can be described by a pair of functions $f, g : \{0,1\}^n \to \{0,1\}^k$ indicating the outputs produced by Alice and Bob, respectively.

In our problem, Alice and Bob need to agree on an input that is uniformly random. We will consider a relaxed scenario where the outputs of Alice and Bob do not need to be uniformly random, but sufficiently close to having high "entropy". To formalize this we introduce some standard definitions.

We recall the *statistical distance* between $\mathcal{D}$ and $\mathcal{D}'$ over sample space Ω is $\sum_{\omega\in\Omega} |\Pr_{\mathcal{D}}(\omega) - \Pr_{\mathcal{D}'}(\omega)|$. We say a distribution $\mathcal{D}$ has *min-entropy* t the probability of every element is at most 2^{-t}. A distribution $\mathcal{D}$ is *δ-close to min-entropy t* if there exists a distribution of min-entropy t which is within statistical distance δ of $\mathcal{D}$. Abusing notation, we will say that a function $f : \{0,1\}^n \to \{0,1\}^k$ has (δ-close to) min-entropy t if the distribution $f(x)$, where x is uniform over $\{0,1\}^n$, has (δ-close to) min-entropy t.

Theorem 1. *For any two functions $f, g : \{0,1\}^n \to \{0,1\}^k$ that are δ-close to min-entropy t and every $\varepsilon \leq 1/2$,*

$$\Pr_{(x,y)_\varepsilon}[f(x) = g(y)] < 2^{-t\epsilon/(1-\epsilon)} + 2\delta.$$

In particular, if the output of Alice and Bob is exactly uniform then $k = t$ and $\delta = 0$, so if they both output $1/\epsilon$ common bits they cannot hope to agree with probability better than $1/2$.

To prove the theorem, we will use the following two well known claims. Claim 2 follows from the fact that $\mathrm{E}_{(x,y)_\varepsilon}[f(x)g(y)]$ is an inner product of f and g. Claim 3 is a corollary of the hypercontractive inequality [Bon70, Bec75] as it is used in [KKL88]. The proofs of these claims require some additional notation. We first show how they imply the theorem.

Claim 2. *For every pair of functions $f, g : \{0,1\}^n \to \mathbb{R}$,*

$$\mathrm{E}_{(x,y)_\varepsilon}[f(x)g(y)] \leq \sqrt{\mathrm{E}_{(x,y)_\varepsilon}[f(x)f(y)]\,\mathrm{E}_{(x,y)_\varepsilon}[g(x)g(y)]}.$$

Claim 3. *For every function $h : \{0,1\}^n \to \{0,1\}$, $\mathrm{E}[h(x)h(y)] \leq \mathrm{E}[h(x)]^{1/(1-\epsilon)}$.*

Proof of Theorem 1. Assume first f and g have min-entropy t. For every $z \in \{0,1\}^k$, let $f_z : \{0,1\}^n \to \{0,1\}$ be the function

$$f_z(x) = \begin{cases} 1, & \text{if } f(x) = z \\ 0, & \text{otherwise.} \end{cases}$$

Define g_z similarly. Then $\mathrm{E}[f_z(x)]$ and $\mathrm{E}[g_z(x)]$ are upper bounded by 2^{-t}. Therefore

$$\begin{aligned}
\Pr[f(x) = g(y)] &= \sum_{z\in\{0,1\}^k} \Pr[f(x) = z \wedge g(y) = z] \\
&= \sum_{z\in\{0,1\}^k} \mathrm{E}[f_z(x)g_z(y)] \\
&\leq \sum_{z\in\{0,1\}^k} \sqrt{\mathrm{E}[f_z(x)f_z(y)] \cdot \mathrm{E}[g_z(x)g_z(y)]} && \text{by Claim 2} \\
&\leq \sum_{z\in\{0,1\}^k} \sqrt{\mathrm{E}[f_z(x)]^{1/(1-\varepsilon)}} \cdot \sqrt{\mathrm{E}[g_z(x)]^{1/(1-\varepsilon)}} && \text{by Claim 3} \\
&\leq \sqrt{\sum_{z\in\{0,1\}^k} \mathrm{E}[f_z(x)]^{1/(1-\varepsilon)}} \cdot \sqrt{\sum_{z\in\{0,1\}^k} \mathrm{E}[g_z(x)]^{1/(1-\varepsilon)}} && \text{by Cauchy-Schwarz}
\end{aligned}$$

Since f and g have min-entropy t it follows that $p_z = \mathrm{E}[f_z(x)] \leq 2^{-t}$ and similarly for g. We can now bound the expression in the first square root by

$$\sum_{z\in\{0,1\}^k} p_z^{1/(1-\varepsilon)} = \sum_{z\in\{0,1\}^k} p_z \times p_z^{\varepsilon/(1-\varepsilon)} \leq 2^{-t\varepsilon/(1-\varepsilon)} \sum_{z\in\{0,1\}^k} p_z = 2^{-t\varepsilon/(1-\varepsilon)}.$$

By an analogous calculation for the second expression, we obtain that $\Pr[f(x) = g(y)] \leq 2^{-t\varepsilon/(1-\varepsilon)}$.

In the case where f and g are δ close min entropy t distributions we proceed as follows. Let $\delta' > \delta$. Then by possibly taking a larger value of n, there exist f' and g' of min entropy t such that such that $\Pr[f \neq f'] \leq \delta'$ and $\Pr[g \neq g'] \leq \delta'$. Now:

$$\Pr[f(x) = g(y)] \leq \Pr[f'(x) = g'(y)] + \Pr[f(x) \neq f'(x)] + \Pr[g(y) \neq g'(y)] \leq 2^{-t\varepsilon/(1-\varepsilon)} + 2\delta'.$$

Since $\delta' > \delta$ is arbitrary the proof follows. □

We now prove the two claims. For this we make use of the Fourier expansion of Boolean functions: Every function $f : \{0,1\}^n \to \mathbb{R}$ can be uniquely written as

$$f(x) = \sum_{S \subseteq [n]} \hat{f}_S \cdot \chi_S(x)$$

where the *character functions* χ_S are given by

$$\chi_S(x) = (-1)^{\sum_{i \in S} x_i}.$$

The characters are orthonormal with respect to the inner product $\langle f, g \rangle = \mathrm{E}[f(x)g(x)]$.

It follows by a calculation that

$$\mathrm{E}_{(x,y)_\varepsilon}[f(x)g(y)] = \sum_{S \subseteq [n]} \hat{f}_S \hat{g}_S \rho^{2|S|} \tag{1}$$

where $\rho = \sqrt{1 - 2\epsilon}$.

Therefore, to prove Claim 2 we observe that

$$\begin{aligned}
\mathrm{E}_{(x,y)_\varepsilon}[f(x)g(y)] &= \sum_{S \subseteq [n]} (\hat{f}_S \rho^{|S|}) \cdot (\hat{g}_S \rho^{|S|}) \\
&\leq \sqrt{\sum_{S \subseteq [n]} \hat{f}_S^2 \rho^{2|S|}} \cdot \sqrt{\sum_{S \subseteq [n]} \hat{g}_S^2 \rho^{2|S|}} && \text{by Cauchy-Schwarz} \\
&= \sqrt{\mathrm{E}_{(x,y)_\varepsilon}[f(x)f(y)]\, \mathrm{E}_{(x,y)_\varepsilon}[g(x)g(y)]}.
\end{aligned}$$

To prove Claim 3, we make use of the hypercontractive inequality [Bon70, Bec75]. This inequality states that for every function $f : \{0,1\}^n \to \mathbb{R}$, we have

$$\mathrm{E}[((T_\rho f)(x))^2]^{1/2} \leq \mathrm{E}[f(x)^{1+\rho^2}]^{1/(1+\rho^2)} \tag{2}$$

where $T_\rho f : \{0,1\}^n \to \mathbb{R}$ is defined via the Fourier expansion of f as the function

$$(T_\rho f)(x) = \sum_{S \subseteq [n]} \hat{f}_S \rho^{|S|} \chi_S(x)$$

Comparing this with (1), we have that

$$\mathrm{E}_x[((T_\rho f)(x))^2] = \mathrm{E}_{(x,y)_\varepsilon}[f(x)f(y)]$$

Where $\rho = \sqrt{1-2\epsilon}$. Now, applying the hypercontractive inequality (2) to a function $h : \{0,1\}^n \to \{0,1\}$ we obtain

$$\mathrm{E}_{(x,y)}[h(x)h(y)]^{1/2} \leq \mathrm{E}[h(x)^{1+\rho^2}]^{1/(1+\rho^2)} = \mathrm{E}[h(x)]^{1/(2-2\epsilon)}$$

which proves Claim 3.

3 A better strategy

We now show that when the agreement probability is sufficiently low, the trivial strategy can be outperformed, and in fact one can get strategies that approach the upper bound from Theorem 1 to within a constant factor.

Theorem 4. *Assume $k \geq 10 + 2(1-\varepsilon)/\varepsilon$, and let $n = n(k,\varepsilon)$ be sufficiently large. There exists a function $f\colon \{0,1\}^n \to \{0,1\}^k$ such that for all $z \in \{0,1\}^k$ it holds that*

$$\forall z \in B^k \ \Pr[f(x) = z] = 2^{-k}, \quad \forall z \in B^k \ \Pr[f(x) = f(y) = z | f(x) = f(y)] = 2^{-k},$$

$$\Pr_{(x,y)_\varepsilon}[f(x) = f(y)] \geq 0.003(\varepsilon k)^{-1/2} 2^{-k\varepsilon/(1-\varepsilon)}$$

The protocol has the following form. Before starting, Alice and Bob agree on a subset C of $\{0,1\}^n$ of size 2^k. On input x (respectively y), Alice (respectively Bob) finds and outputs the index of the closest point in C (with an explicit rule in case of ties). We will show that there exists a choice of C for which (1) each output is generated with the same probability and (2) the probability of agreement is high.

In fact, we prove that on average, a random subspace of $\{0,1\}^n$ of dimension k has both properties (1) and (2). In our analysis, we fix k and the noise ϵ and let n go to infinity.

Let C be an affine subspace of $\{0,1\}^n$. Write $C = a + L$ where L is a linear subspace. Let $\prec$ define a strict total order on $\{0,1\}^n$ with the property that if the Hamming weight of x is smaller than the Hamming weight of y then $x \prec y$. We define the regions $R_c, c \in C$ by:

$$R_c = \{x\colon x + c \prec x + c' \text{ for all } c \neq c' \in C\}$$

Note that if c is the unique closest point to x among all the points in C then $x \in R_c$.

Let $H : L \to \{0,1\}^k$ be any invertible linear map and let $f : \{0,1\}^n \to \{0,1\}^k$ be defined as $f(x) = H(c)$, where c is the unique point such that $x \in R_{a+c}$.

Claim 5. *For all $z \in \{0,1\}^k$,*

$$\Pr[f(x) = z] = 2^{-k} \quad \textit{and} \quad \Pr[f(x) = f(y) = z | f(x) = f(y)] = 2^{-k}.$$

Proof. If $a+c, a+c' \in C$ and $x+c \in R_{a+c}$ then $x+c' \in R_{a+c'}$. So for every $c \in L$ we have $f(x+c) = f(x) + H(c)$. Let $z, z' \in \{0,1\}^k$ and let $z' = z + H(c)$ where $c \in L$. Then (x, y) and $(x+c, y+c)$ have the same distribution and therefore

$$\Pr[f(x) = f(y) = z'] = \Pr[f(x+c) = f(y+c) = z + H(c)] = \Pr[f(x) = f(y) = z],$$

and similarly

$$\Pr[f(x) = z'] = \Pr[f(x) = z + H(c)] = \Pr[f(x+c) = z] = \Pr[f(x) = z],$$

as needed. □

Let t be chosen so that

$$\frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-z^2/2} dz = 2^{-k-2}$$

Let C be a random affine space in $\{0,1\}^n$ of dimension k. Let $r = n/2 + t\sqrt{n}/2$ and note that by the central limit theorem the hamming ball of radius r contains $2^{n-k-1}(1-o(1))$ points as $n \to \infty$. We will say $x \in \{0,1\}^n$ is *covered* by $c \in C$ (denoted by $x \in B_c$) if x belongs to the ball of radius r centered at c. We say x is *uniquely covered* by c (denoted by $x \in U_c$) if it is covered by c but not by any other $c' \in C$. Observe that $U_c \subseteq R_c$.

Claim 6. *Let C be a random affine subspace of $\{0,1\}^n$ of dimension k. Then for n sufficiently large,*

$$\mathrm{E}_C \Pr_{(x,y)_\varepsilon}[\exists c \in C \colon x, y \in U_c] \geq \frac{1}{8} \cdot \Pr[Z > \sqrt{\varepsilon/(1-\varepsilon)}t]$$

where Z is a normal variable of mean 0 and variance 1.

By Claims 5 and 6, there must exist a set of points C for which (1) all the regions R_c and of the same size and (2) $\Pr_{(x,y)_\varepsilon}[x, y \in R_c \text{ for some } c \in C] \geq \frac{1}{8} \cdot \Pr[Z > \sqrt{\varepsilon/(1-\varepsilon)}t]$. To finish the proof of Theorem 4 we calculate a lower bound for the last expression.

Claim 7. *Let $k \geq 10 + 2(1-\varepsilon)/\varepsilon$. Then*

$$\frac{1}{8} \cdot \Pr[Z > \sqrt{\varepsilon/(1-\varepsilon)}t] \geq 0.003(\varepsilon k)^{-1/2} 2^{-\varepsilon k/(1-\varepsilon)}$$

where Z is a normal variable of mean 0 and variance 1.

Proof. We will use the following estimates valid for every $y > 0$:

$$\frac{y}{y^2+1} e^{-y^2/2} \leq \int_y^\infty e^{-z^2/2} dz \leq \frac{1}{y} e^{-y^2/2}.$$

Note that if $k \geq 10$ then $t \geq 3$ and therefore

$$2^{-k-2} = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-z^2/2} dz \leq \frac{1}{\sqrt{2\pi}t} e^{-t^2/2} \leq e^{-t^2/2}$$

which implies that $t \le \sqrt{2(k+2)\ln 2} \le \sqrt{2k}$. Moreover,

$$2^{-k} \ge 2^{-k-2} = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-z^2/2} dz \ge \frac{1}{\sqrt{2\pi}} \frac{t}{t^2+1} e^{-t^2/2} \ge \frac{1}{\sqrt{2\pi} 2t} e^{-t^2/2} \ge e^{-t^2},$$

which implies that $t \ge \sqrt{k \ln 2}$. So if $k \ge 10 + 2(1-\varepsilon)/\varepsilon$, then $t \ge \sqrt{(1-\varepsilon)/\varepsilon}$ and therefore

$$\begin{aligned}
\Pr[Z > \sqrt{\varepsilon/(1-\varepsilon)}t] &> \frac{1}{\sqrt{2\pi}} \cdot \frac{\sqrt{\varepsilon/(1-\varepsilon)}t}{(\varepsilon/(1-\varepsilon))t^2+1} \cdot e^{-\varepsilon t^2/2(1-\varepsilon)} \\
&\ge \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{2\sqrt{\varepsilon/(1-\varepsilon)}t} \cdot \left(e^{-t^2/2}\right)^{\varepsilon/(1-\varepsilon)} \\
&\ge \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{2\sqrt{2\varepsilon/(1-\varepsilon)k}} 2^{-\varepsilon(k+2)/(1-\varepsilon)} \\
&\ge 0.024 \cdot (\varepsilon k)^{-1/2} 2^{-\varepsilon k/(1-\varepsilon)}
\end{aligned}$$

as needed. □

Proof of Claim 6. Let C be a random affine k-dimensional space of $\{0,1\}^n$. Such a space can be constructed by starting with a random point $c_0 \sim \{0,1\}^n$, and iteratively constructing the space $C_i = c_0 + \mathrm{span}(c_0 + c_1, \ldots, c_0 + c_i)$, where c_i is chosen uniformly from $\{0,1\}^n \setminus C_{i-1}$. Finally let $C = C_k$. From the construction of C it follows that $AC + c$ has the same distribution as C for every invertible linear transformation A and every vector c. Since for every pair of vectors $a \ne a', b \ne b'$ there exsits an invertible A and a c such that $Aa = b$ and $Aa' = b'$ it follows that $\Pr_C[a, a' \in C] = \Pr_C[b, b' \in C]$. Then

$$\begin{aligned}
&\mathrm{E}_C \Pr_{(x,y)_\varepsilon}[\exists c \in C \colon x, y \in U_c] \\
&= \mathrm{E}_C \sum_{c \in C} \Pr_{(x,y)_\varepsilon}[x, y \in U_c] \\
&= \mathrm{E}_C \sum_{c \in C} \Pr_{(x,y)_\varepsilon}[x, y \in B_c] \Pr_{(x,y)_\varepsilon}[\forall c' \ne c \colon x, y \notin B_{c'} \mid x, y \in B_c] \\
&\ge \mathrm{E}_C \sum_{c \in C} \Pr_{(x,y)_\varepsilon}[x, y \in B_c] \Big(1 - \sum_{c' \ne c} \Pr_{(x,y)_\varepsilon}[x \in B_{c'} \text{ or } y \in B_{c'} \mid x, y \in B_c]\Big) \\
&= \sum_{\{0,1\}^k} \mathrm{E}_{a \sim \{0,1\}^n} \Big[\Pr[x, y \in B_a] \Big(1 - \sum_{a' \ne a} \mathrm{E}_{a' \sim \{0,1\}^n \setminus \{a\}} \Pr[x \in B_{a'} \text{ or } y \in B_{a'} \mid x, y \in B_a]\Big)\Big].
\end{aligned}$$

The last line uses the fact that the distribution over any pair of points $c \ne c'$ in a random affine space (of dimension at least 1) is the same as the uniform distribution over pairs $a, a' \in \{0,1\}^n$ conditioned on $a' \ne a$. For the expression in the inner summation, we have

$$\begin{aligned}
&\mathrm{E}_{a' \sim \{0,1\}^n} \Pr_{(x,y)_\varepsilon}[x \in B_{a'} \text{ or } y \in B_{a'} \mid x, y \in B_a] \\
&\qquad \le 2 \,\mathrm{E}_{a' \sim \{0,1\}^n} \Pr_{(x,y)_\varepsilon}[x \in B_{a'} \mid x, y \in B_a] = 2 \Pr_x[x \in B_0] \le 2^{-k-1}
\end{aligned}$$

and therefore

$$\mathrm{E}_{a' \sim \{0,1\}^n \setminus \{a\}} \Pr_{(x,y)_\varepsilon}[x, y \in B_{a'} \mid x, y \in B_a] \leq 2^{-k-1} \frac{2^n}{2^n - 1}.$$

from where the desired expression equals at least

$$\sum_{\{0,1\}^k} \mathrm{E}_a \Pr_{(x,y)_\varepsilon}[x, y \in B_a] \cdot (1 - (2^k - 1)2^{-k-1} \frac{2^n}{2^n - 1}) > 2^k \cdot \mathrm{E}_a \Pr_{(x,y)_\varepsilon}[x, y \in B_a] \cdot (1/2)$$
$$= 2^{k-1} \cdot \Pr_{(x,y)_\varepsilon}[x, y \in B_0].$$

To calculate the last expression, by the two-dimensional central limit theorem we have

$$\Pr_{(x,y)_\varepsilon}[x, y \in B_0] \to \Pr_{X,Z}[X > t, \theta X + \sqrt{1 - \theta^2} Z > t] \qquad \text{as} \qquad n \to \infty$$

where $\theta = 1 - 2\varepsilon$ and X, Z are independent normal variables with mean 0 and variance 1. We now lower bound this expression:

$$\begin{aligned} \Pr_{X,Z}[X > t, \theta X + \sqrt{1 - \theta^2} Z > t] &= \Pr[X > t] \Pr[\theta X + \sqrt{1 - \theta^2} Z > t \mid X > t] \\ &\geq \Pr[X > t] \Pr[\theta t + \sqrt{1 - \theta^2} Z > t] \\ &= \Pr[X > t] \Pr[Z > \sqrt{\varepsilon/(1 - \varepsilon)} t]. \end{aligned}$$

Recalling that as $n \to \infty$, $\Pr[X > t] \to 2^{-k-2}$, we obtain that as n becomes sufficiently large,

$$\mathrm{E}_C \Pr_{(x,y)_\varepsilon}[x, y \in U_c \text{ for some } c \in C] \geq \frac{1}{8} \cdot \Pr[Z > \sqrt{\varepsilon/(1 + \varepsilon)} t]. \qquad \square$$

4 Conclusion

In this work we propose the following protocol for two parties that are given access to n noisy random bits with noise of rate ε to agree on a common random string of length k:

Preprocessing stage:

1. Define a strict total order $\prec$ on $\{0,1\}^n$ that is consistent with the partial order induced by Hamming weight.
2. Choose a random k-dimensional affine subspace C of $\{0,1\}^n$. Identify the elements of C with strings in $\{0,1\}^k$.

Decoding stage: On input x, output the unique $c \in C$ such that $x + c \prec x + c'$ for all $c' \in C$, $c' \neq c$.

Our analysis shows that on average over the choice of C, the outputs of Alice and Bob agree with probability $\Omega((k\varepsilon)^{-1/2} 2^{-k\varepsilon/(1-\varepsilon)})$, which is best possible up to a factor of $O(\sqrt{k\varepsilon})$ provided that $k \geq 2/\varepsilon + O(1)$ and $n = n(k, \varepsilon)$ is sufficiently large.

We remark that an explicit upper bound on n in terms of k and ε can in principle be obtained by using a quantitative version of the central limit theorem in our arguments.

We leave open the question of designing a deterministic and more efficient protocol for the problem considered here. It may also be interesting to investigate how much common randomness can be extracted from other noisy channels (x, y).

Acknowledgments

We are grateful to Philip Leong for explaining the problem of circuit unique ID extraction which served as the original motivation for the paper and to Uri Feige for crucial insights used in our construction.

References

[AC93] R. Ahlswede and I. Csiszàr. Common randomness in information theory and cryptography – part i: Secret sharing. *IEEE Trans. Inform. Theory*, 39(4):1121–1132, July 1993. 2

[Bec75] W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, (102):159–182, 1975. 4, 5

[Bon70] A. Bonami. Etude des coefficients de Fourier des fonctions de $l^p(g)$. *Annales de l'institut Fourier*, 20(2):335–402, 1970. 4, 5

[GK72] P. Gács and J. Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):119–162, 1972. 2

[KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *FOCS '88: Proceedings of the 29th Annual Symposium on Foundations of Computer Science*, pages 68–80, Washington, DC, USA, 1988. IEEE Computer Society. 4

[LLG+05] D. Lim, J. W. Lee, B. Gassend, G. E. Suh, M. van Dijk, and S. Devadas. Extracting Secret Keys From Integrated Circuits. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 13(10):1200–1205, October 2005. 1

[Mau93] U. Maurer. Secret key agreement by public discussion based on common information. *IEEE Trans. Inform. Theory*, 39:733–742, May 1993. 2

[MO05] E. Mossel and R. O'Donnell. Coin flipping from a cosmic source: On error correction of truly random bits. *Random Structures & Algorithms*, 4(26):418–436, 2005. 2

[MOR+06] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif, and B. Sudakov. Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality. *Israel J. Math.*, 154:299–336, 2006. 2

[SD07] G. Edward Suh and Srinivas Devadas. Physical unclonable functions for device authentication and secret key generation. In *DAC '07: Proceedings of the 44th annual conference on Design automation*, pages 9–14, New York, NY, USA, 2007. ACM. 1

[SHO08] Ying Su, J. Holleman, and B.P. Otis. A digital 1.6 pJ/bit chip identification circuit using process variations. *Solid-State Circuits, IEEE Journal of*, 43(1):69–77, Jan. 2008. 1

[Wit75] H. S. Witsenhausen. On sequences of pairs of dependent random variables. *SIAM Journal on Applied Mathematics*, 28(1):100–113, 1975. 2

[Yan07] K. Yang. On the (im)possibility of non-interactive correlation distillation. *Theoretical Computer Science*, 382(2):157–166, 2007. 2

[YLH+09] Haile Yu, Philip H. W. Leong, Heiko Hinkelmann, Leandro Möller, Manfred Glesner, and Peter Zipf. Towards a unique fpga-based identification circuit using process variations. In *19th International Conference on Field Programmable Logic and Applications, FPL 2009, August 31 - September 2, 2009, Prague, Czech Republic*, pages 397–402, 2009. 1